Canadian Labour Market and Skills Researcher Network

Working Paper No. 49

Selecting Economic Immigrants: An Actuarial Approach

John McHale
Queen's University

Keith Rogers
Queen's University

November 2009

CLSRN is supported by Human Resources and Social Development Canada (HRSDC) and the Social Sciences and Humanities Research Council of Canada (SSHRC).
All opinions are those of the authors and do not reflect the views of HRSDC or the SSHRC.

Selecting Economic Immigrants: An Actuarial Approach

John McHale and Keith Rogers[1]

Queen's School of Business

Queen's University

jmchale@business.queensu.ca

krogers@business.queensu.ca

[1] We would like to thank Chona Iturralde, Martha Justus, Frances Koo, Stanley Kustec, Michael-John McCormick, Elizabeth Ruddick, Eden Thompson and Sebastien Vachon of Citizenship and Immigration Canada for providing their insights into the immigrant selection process and for their facilitation role in accessing the IMDB; Mike Abbott and Charles Beach at Queen's Economics Department for their help with understanding the IMDB; and the Canadian Labour Market and Skills Researcher Network for its financial support and valuable feedback. We owe a special thanks to Tristan Cayn from CIC who worked tirelessly with us to run our code at Statistics Canada.

Abstract

There is growing international interest in a Canadian-style points system for selecting economic immigrants. Although existing points systems have been influenced by the human capital literature, the findings have traditionally been incorporated in an ad hoc way. This paper explores a formal method for designing a points system based on a human capital earnings regression for predicting immigrant economic success. The method is implemented for Canada using the IMDB, a remarkable longitudinal database that combines information on immigrants' characteristics at landing with their subsequent income performance as reported on tax returns. We demonstrate the feasibility of the method by developing an illustrative points system. We also explore how the selection system can be improved by incorporating additional information such as country-of-origin characteristics and intended occupations. We discuss what our findings imply for the debate about the relative merits of points- and employment-based systems for selecting economic immigrants.

JEL Code: F22, J24, J31

Keywords: Immigrant selection, Points system, Human capital, Earnings prediction

1. Introduction

There is growing international interest in a Canadian-style points system for selecting economic immigrants.[2] At the same time, there is rising concern in Canada about the income performance of recent cohorts of economic immigrants, as many of those selected through the points system struggle in the labour market (Aydemir and Skuterud, 2005; Picot, Hou, and Coulombe, 2007). In this paper, we explore a new approach to the design and evaluation of a points-based selection system. The basic idea is that the system is designed based on the human capital earnings regression that best predicts the earnings of immigrants. We apply this approach to the design of a points system for Canada using the Longitudinal Immigrant Database (IMDB) to develop the prediction model. The IMDB is a remarkable dataset that combines information on the human capital characteristics of immigrants at landing with income data derived from post-landing tax filings, and is uniquely suited to developing our design approach.

The design of the existing points system has undoubtedly been influenced by the vast empirical literature relating immigrant characteristics at landing to their subsequent economic performance. One indication is that the measured human capital—most notably educational attainment—of immigrants admitted under the points system has increased dramatically (see, e.g., Beach, Green, and Worsick, 2006; Picot and Sweetman, 2005). But the design process has followed what can fairly be called a *clinical* rather than an *actuarial* approach: that is, it has depended on expert judgment rather than an explicit statistically-based design.[3] With the multiple objectives that are weighed in any

[2] Points systems are also used for skills-based selection in Australia and New Zealand. The United Kingdom is in the process of introducing a permanent system to replace its points-based Highly Skilled Migrant Programme that was first introduced on a pilot basis in 2002 (United Kingdom Home Office, 2006). Points systems are under consideration in France, Ireland and Spain. In Germany, a points system went down to a narrow legislative defeat in 2003. A points system was not part of the comprehensive immigration reform passed by the U.S. Senate in May of 2006, although it was the subject of hearings of the Senate Committee on Health, Education, Labor, and Pensions in September (see Beach, 2006).

[3] The clinical-actuarial distinction is common in psychology, jurisprudence and medicine. A large literature has developed following Meehl (1954) that compares the predictive success of the two methods. The actuarial method has generally found to be superior where the two methods have access to the same information (see, e.g., Dawes, Faust, and Meehl, 1989). Of course, where an expert (i.e. clinician) has access to information unavailable to someone using the statistical model, it is quite possible that the former will make more accurate predictions. For example, a visa officer interviewing an applicant could make a judgment about the individual's social skills and ambition, information that would not be available to the statistical model. However, the use of this type of information is not what is at issue in the design of a system that is dependent on objectively verifiable information. The design problem relates to how best to weigh the various available pieces of information (educational attainment, fluency in official languages,

selection system (economic, fiscal, humanitarian, family reunification, etc.), it is inevitable that judgment is applied in the system's overall design. However, we think an approach that focuses directly on predicted earnings is both appropriate and feasible for the economic immigrant stream, where the objective is selecting high-earning immigrants that will strengthen the economy and fiscal system for the benefit of the pre-immigration population.[4]

We thus develop an actuarial—or optimal-prediction-based-on-historical-data—approach to the design of the system for selecting economic immigrants. The central idea is to use data on the landing characteristics and subsequent income performance of earlier immigrant cohorts to identify the "best" human capital-based prediction equation. This equation is combined with an explicit threshold for predicted earnings below which applicants are not accepted. The point allocations are then objectively mapped from the parameters of this prediction equation given the chosen threshold. We also develop the concept of the selection frontier as a means for evaluating selection systems. The frontier shows the tradeoff between the expected earnings of the pool of admitted immigrants and the number of immigrants admitted, with each point on the frontier mapping to a unique predicted earnings threshold. The position of the frontier will depend on predictive success of the underlying earnings regression, with the best prediction equation leading to the highest feasible frontier.

Although there is growing interest in points systems among industrial countries, there is also a debate about their effectiveness. One view is that the quality of a country's immigrant stream is dominated by the pool of people who desire to move to the country (Jasso and Rosenzweig, 2005). This suggests that fine-tuning the selection

---

etc.) In this equal-information setting, it is harder to see an advantage for expert judgment over statistical models that are chosen to best fit the historical data.

[4] Even from a narrow economic perspective of those already present in the host country, a better measure of economic value is the "surplus" that the country gains from the immigrants. This surplus can be defined as the value the country receives less what they must pay to the immigrants. Simple models show that it is not necessarily the most highly skilled immigrants that generate the greatest surplus (see, e.g., Borjas, 1995). However, the relevance of human capital is likely to increase when we allow for fiscal effects, knowledge spillovers, or the value of specialized skills. Augmenting the relative supply of skilled workers should also reduce overall earnings inequality, so that skilled recruitment can be desirable on both efficiency and equity grounds. But whatever the merits of focusing narrowly on skills, it is the case that a number of countries are striving to select more skilled and higher earning immigrant pools. It is thus worthwhile to look for a more systematic approach to designing a skills-based selection system.

system is unlikely to have significant effects. Another view holds that the design of the selection system does have first order effects on immigrant labour market success (see, e.g., Lester and Richardson on the comparison of the Canadian and Australian points systems). The actuarial approach allows us to examine the potential for fine-tuning a conventional points system, say by adjusting allocations to the usual point sources such as education, experience and language skills. Subject to data availability, this approach also allows us explore, in a systematic, way the potential contribution of less conventional sources of points that have been suggested by the human-capital literature (country-of-origin, achievement on literacy tests, quality of educational institution, pre-emigration earnings, etc.). By exploring the performance of the best designed point systems, the actuarial approach should help inform whether there is a need for more radical departures from points-based selection, such as strict pre-immigration job offer requirements or probationary periods on temporary work visas before permanent immigration status is granted.

The rest of the paper is structured as follows. In the next section, we review the related literature on immigrant assimilation and the effectiveness of selection systems. We then describe our design methodology in Section 3 and our data in Section 4. Section 5 then develops an illustrative points system and discusses various extensions. Section 6 concludes with a discussion of what our results imply about the effectiveness of even the best designed points system.

2. Related literature

Following Chiswick (1978), a large literature has developed that explores how human capital characteristics at landing affect an immigrant's subsequent labour market success. A major theme in this literature is how more recent immigrant cohorts to the United States compare with earlier cohorts in terms of entry earnings and subsequent earnings growth (e.g. Borjas, 1985; Duleep and Regets, 2002). A substantial parallel literature has developed looking at Canadian immigrants. A central focus has been the declining performance of more recent immigrant cohorts relative to native-born workers (Abbot and Beach, 1993; Baker and Benjamin, 1994; Bloom, Grenier, and Gunderson, 1995; Grant, 1999; Frenette and Morissette, 2003; Green and Worswick, 2004; Aydemir and Skuterud, 2005; Picot, Hou, and Coulombe; 2007). Two important findings have been that immigrants that come at young ages tend to perform better (possibly reflecting the acquisition of Canadian schooling) and that there is a negligible return to foreign experience.[5] A number of recent studies have looked at how less traditional human capital measures are associated with labour market performance: credential acquisition or "sheepskin effects" (Ferrer and Riddel, 2004); source-country educational quality (Sweetman, 2004); and literacy skills (Ferrer, Green, and Riddell, 2004; Alboim, Finnie, and Meng, 2005).

There is also a smaller literature that looks at how alternative immigrant selection systems affect immigrant characteristics and performance. A key issue in this literature is whether immigrant quality is affected more by who desires to emigrate to a particular host country or by the selection system that the host country employs. Jasso and Rosenzweig (1995) find a small difference between the performance of U.S. immigrants screened for skills and those who gain admission based on family ties. More recently, Jasso and Rosenzweig (2005) find little difference in the operation of the employment-based U.S. and the skills-based Australian selection systems, leading them to conclude that the immigrant mix is largely driven by the self-selection decisions of the immigrants. Antecol, Cobb-Clark, and Trejo (2001) find that immigrants to Australia and Canada do have more measured human capital than immigrants to the U.S., but conclude that this has more to do with the latter's geographic and historic ties to Mexico than with differences in selection systems.

For Canada, Beach, Green, and Worswick (2006) have found variations in the Canadian selection system—including variations in the way points are allocated for different measures of human capital—do impact the characteristics of the admitted immigrants. Adyemir (2002) develops and empirically implements a model that allows for both self-selection and host-country selection, and finds that both are important in determining who actually immigrates. In a direct comparison of the Australian and Canadian points-based systems, Lester and Richardson (2004) argue that reforms to the

[5] See Schaafsma and Sweetman (2001). See also Friedberg (2002) for a similar finding of a low return to foreign experience for immigrants to Israel.

Australian system explain the recent better performance of the Australian economic immigrants (see also Hawthorne, 2005).

It is difficult to sum up these substantial literatures. The immigrant-assimilation literature has certainly demonstrated the usefulness of human capital-based earnings regressions for predicting immigrant success. The selection-system literature shows that when a system selects for given characteristics it does tend to have an immigration flow with those characteristics, but there is less agreement on how much the selection system can affect the labour market success of the admitted immigrants. As far as we can tell, there has not been a previous attempt to explore the optimal design of the selection system using human capital-based earnings prediction. With the proposed actuarial approach, we hope to identify the optimal selection in a rigorous and transparent way, to explore the scope for fine-tuning a conventional points system to improve immigrant selection, and to contribute to the debate about the merits of points-based selection.

3. Optimal Design Methodology

In this section, we summarize the basic steps for identifying the optimal points system. (This method builds on McHale and Rogers (2007) to which the reader is referred for additional details.) The inputs for this method are a human capital-based earnings regression for making predictions of immigrant labor-market success and a (lifetime) predicted-earnings threshold for deciding who to accept. The outputs are the point allocations per unit of each human capital characteristic.

To illustrate the basic design approach, we assume that an immigrant's earnings depend only on their years of schooling (S_i), years of experience (E_i), and years since migration (t_i). Host-country earnings are given by a standard log-linear earnings regression:

(1) $$\ln Y_{it} = y_{it} = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 t_i + u_{it} \qquad u_{it} \sim n(0, \sigma_u^2) .$$

We use regression analysis to obtain a predictor of log earnings,

(2) $$\hat{y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 E_i + \hat{\beta}_3 t_i$$

To obtain the log of expected earnings, we use the approximation,

$$(3) \qquad \ln \hat{Y}_{it} \approx \frac{\hat{\sigma}_u^2}{2} + \hat{y}_{it} = \frac{\hat{\sigma}_u^2}{2} + \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 E_i + \hat{\beta}_3 t_i .$$

Letting T_i represent the number of years that the immigrant will be in the host labour market and letting δ represent the discount rate, we can use (3) to write the present discounted value of predicted earnings as,[6]

$$(4) \qquad \begin{aligned} \hat{Z}_i &= \int_0^{T_i} e^{-\delta t_i} \hat{Y}_{it} dt_i \\ &= e^{\frac{\hat{\sigma}_u^2}{2} + \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 E_i} \int_0^{T_i} e^{(\hat{\beta}_3 - \delta) t_i} dt_i \\ &= e^{\frac{\hat{\sigma}_u^2}{2} + \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 E_i} \left(\frac{1}{\hat{\beta}_3 - \delta} \right) \left(e^{(\hat{\beta}_3 - \delta) T_i} - 1 \right) \end{aligned}$$

We assume that the immigrant will work until age $\overline{A}$, so that $T_i = \overline{A} - A_i$, where A_i is age at landing. Making this substitution and taking logs yields,

$$(5) \qquad \ln \hat{Z}_i = \frac{\hat{\sigma}_u^2}{2} + \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 E_i - \ln\left(\hat{\beta}_3 - \delta\right) + \ln\left(e^{(\hat{\beta}_3 - \delta)(\overline{A} - A_i)} - 1\right)$$

We next assume that the policy maker sets a threshold, $\hat{Z}^*$, for predicted lifetime earnings. Any applicant with predicted earnings at or above this level is accepted; others rejected. To obtain the points allocations, we arbitrarily set the point allocation to 100 for someone with predicted earnings that is exactly equal to the threshold. With

[6] Assuming that there is no out-migration of emigrants, δ will reflect the rate at which the policy maker discounts future earnings relative to current earnings. However, assuming a constant conditional probability of exit (or "hazard rate"), δ can also conveniently include a discount due to expected attrition due to out-migration. [Making use of our longitudinal dataset, we plan to explore the estimation of such hazard rates in the next version of the paper.]

predicted earnings set equal to the threshold, we can rearrange equation (5) to obtain the per unit points allocation for each of the human capital characteristics.

$$\frac{100\tilde{\beta}}{\tilde{\beta}} = \left(\frac{100\hat{\beta}_1}{\tilde{\beta}}\right)S_i + \left(\frac{100\hat{\beta}_2}{\tilde{\beta}}\right)E_i + \left(\frac{100\ln\left(e^{(\hat{\beta}_3-\delta)(\bar{A}-A_i)}-1\right)}{\tilde{\beta}}\right),$$

(6)

$$\text{where} \quad \tilde{\beta} = \ln\hat{Z}^* - \frac{\hat{\sigma}_u^2}{2} - \hat{\beta}_0 + \ln\left(\hat{\beta}_3 - \delta\right).$$

The respective point allocations for schooling and experience are given by the relevant terms in parentheses in equation (6). The last term in the equation determines the non-linear point allocations given for age-at landing. Based on these allocations, any applicant that scores 100 points or above has above-threshold predicted earnings and is accepted.

How do we know if the selection system is performing well? In Appendix 1, we describe an evaluation tool that we call the selection frontier. This frontier shows the trade-off between the "quality" of the selected immigrants as measured by the expected value of the admitted pool and the "quantity" (or number) of immigrants admitted. Each point on the frontier is shown is map to a unique predicted earnings threshold. We show in the appendix that improvements in the explanatory power of the regression (conveniently measured by the regression R^2) are associated with upward shifts of the frontier—i .e. a higher quality pool for any number of immigrants admitted. The frontier provides a useful tool to judge the effectiveness of a given points system and to explore the gains from fine tuning the point allocations.

4. Data[7]

The actuarial approach to designing a points system depends on the availability of historical data on both the characteristics of immigrants at landing and their subsequent performance in the labour market. As mentioned above, the Canadian IMDB is an ideal

[7] The information in this section draws heavily on publicly-available documentation of the IMDB provided by Citizenship and Immigration Canada and Statistics Canada as well as Abbott (2003) and personal communication with CIC employees.

source of both types of data. The IMDB is an administrative database containing information on immigrants to Canada since 1980. It combines static information from an immigrant's landing records with tax earnings with income data from tax filings.[8] It is worth noting that this is not a sample of immigrants, but rather the population of immigrants with at least one personal income tax filing. The database is updated annually as new immigrant cohorts arrive and new tax data becomes available. Tax return data is only recorded for the fifteen years after the first tax filing, so that we have at most 15 annual income observations on each immigrant.

In terms of static (or "tombstone") data on landing characteristics, it contains basic demographic data (sex, age, country of birth); skill measures (English / French language ability, native language, years of schooling, educational attainment); intended settlement in Canada (province and city, industry and occupation); family status; and admission details (immigrant category, applicability of points system, allocation of sufficient points for admission, principal applicant flag.) The dynamic data consists of up to 15 years of income data by income type (e.g. employment earnings, investment income, rental income etc.)

Our interest in this data is principally centered on the economic immigrants as opposed to those admitted under the family class or refugees. We are further interested in separating the impact of earnings from returns to capital so we exclude immigrants from the investor and entrepreneur classes. The selection criteria then are: principal applicants between the ages of 18 and 64 who enter in IMCAT category 7 (skilled workers principal applicant abroad no special program) or IMCAT category 8 (skilled workers principal applicant in Canada or with special program). For this group our dependent variable is the log of total earned income, which is the sum of all reported earnings on the individual's tax records.

The IMDB does have some limitations. First, a number of immigrants to Canada have never filed a tax return and therefore do not appear in the sample. Second, immigrants can be temporarily or permanently absent from the through return-migration,

[8] Public access to the IMDB is strictly constrained given the sensitive nature of the underlying tax and personal information. As a consequence, we cannot have direct access to the data. We are extremely grateful to Citizenship and Immigration Canada and Statistics Canada for generously agreeing to work with us to implement the required data runs.

on-migration to a third country, or death. This results in an unbalanced panel sample. Third, the IMDB does not contain reason-codes for individuals not filing a tax return so that we can not be sure why an individual has disappeared from the sample.

We have supplemented the IMBD data with aggregate variables for both country-of-destination and country-of origin. Two aggregate Canadian variables are added to control for macroeconomic effects: the national unemployment rate (CANSIM Table 282-0002) and the log of average real annual earnings for full time, full year workers (CANSIM Table 202-0101). Three aggregate country-of-origin variables are included to control for conditions in the source country that might affect the unobserved human capital of immigrants: the log of real GDP per capita adjusted for purchasing power parity (World Bank, World Development Indicators); the log of distance from Canada (taken from Andrew Rose's bilateral trade database[9]); and the 1980-2004 average of the Gini coefficient[10] (World Bank, World Development Indicators). Summary statistics for all variables used in the regressions are shown in Table 1.

5. Empirical Implementation

5.1 Base Regression

To determine the feasibility of the proposed design approach, we first demonstrate the development of a simple points system that is linear in experience, language ability and educational attainment, and non-linear in age-at-landing. The base regression from which this illustrative points system is derived is shown in Table 2. We record two specifications, the first with and the second without an age-at-landing variable.[11] We also record two estimation methods for each specification: ordinary least squares (with

[9] Available at http://faculty.haas.berkeley.edu/arose/.

[10] The Gini coefficient is only available in household survey years. These years vary from country to country.

[11] Recall that age-at-landing still matters for points even if an age-at-landing variable is not included in the regression. The reason is that age-at-landing determines potential years in the Canadian labour market. Although age-at-landing is likely to affect an immigrant's capacity to adapt to the Canadian labour market, we are concerned about our ability to separately identify the age-at-landing and experience effects. This stems from our relatively crude measure of experience: Age-at-Landing – Years of Schooling – 5. Controlling for age-at-landing, the experience effect is then identified by variation in years of schooling (holding educational attainment constant), which we think is a thin reed on which to base identification. We thus concentrate on the results without the age-at-landing variable in developing the illustrative points system.

standard errors that are robust to individual immigrant-level clustering); and random effects to explicitly take account of serial correlation in individual earnings over time.[12]

The dependent variable is the log of real annual earnings expressed in constant 2005 dollars. We impose the additional restriction that annual earnings are greater than $1,000. We do not include earnings observations for the year of landing, since the length of time will typically be less than a full year and will vary across immigrants. Experience is defined as Age at Landing – Years of Schooling – 5. Language enters as a pair of dummy variables: an English dummy that takes the value 1 if English is the immigrant's native language; and a French dummy that takes the value 1 if French is the native language. Educational attainment enters as a set of seven dummy variables (with Primary the excluded category): Secondary, Some Post-Secondary, Trade Certificate, Diploma, Bachelors, Masters, and PhD. We also include a full set of cohort year dummies for the years 1981 to 2003 (with 1980 chosen as the excluded cohort).[13] We include two variables to control for macro/time effects: the national unemployment rate to control for business cycle effects and trend movements in the underlying structural rate of unemployment; and the log of real annual earnings for full time, full year workers to control for secular trends in economy-wide earnings.[14]

[12] We use random effects rather than fixed effects for two reasons. First, under fixed effects, the coefficients on all linear time-invariant explanatory variables cannot be estimated. In our regressions, most of the central variables of interest take this form. And second, given that our interest is in predicting earnings, we are not concerned about correlation between the explanatory variables and the error term (a problem that fixed effects can help fix). Provided that these correlations are stable over time—e.g. high educational attainment is stably correlation with unobserved natural abilities that positively affect earnings—it is advantageous for earnings predictions to be able to use observed human capital characteristics as indicators of unobserved abilities. We thus do not present our estimated coefficients as estimates of the returns to human capital, but rather as associations in the historic data.

[13] There is no 2004 cohort since we include earnings observations only on the first full year after year of landing.

[14] The separation of cohort, years-since-migration and macro effects has been a major focus of the empirical literature on immigrant earnings. Since our data from the IMBD is for immigrants only, we could not utilize the common identification practice of assuming that marco effects are equal for immigrants and natives. We did explore the method originated independently by Hall (1971) and Mason et al. (1973) of including a full set of both cohort and time dummies (in addition to the year-since-migration variable), and imposing the identification constraint that either two of the cohort dummies or two of the time dummies have equal coefficients. As is well-described by Glenn (2005), the results can be highly sensitive to the chosen identification constraint. Some experimentation with alternative conditions showed this to be the case with our data. Thus, lacking a priori grounds for choosing a restriction, we decided against this approach. We caution, however, that the type of cohort, years-since-migration, macro decomposition we are seeking is unavoidably problematic, and additional studies are needed to confirm the robustness of our results.

An obvious concern with our design approach is that the immigrant earnings generation process might not be stable over time. One reason to worry about instability is that past immigrants are obviously a selected sample, and the nature of that selection may change over time. We have tried to minimize the instability in two ways. First, we have limited our sample to immigrants who enter as principal applicants in the skilled worker stream. Since these individuals have been selected through the points system, they have been selected based on observed human capital characteristics. With selection on observables, even the fact that our historic sample is in an obvious way a "selected sample" should not lead to a bias that is sensitive to the precise nature of the selection process. Second, although the composition of the immigrant pool has certainly been changing over time (most obviously in the country-of-origin distribution of the admitted immigrants), we can crudely capture the changing distribution with cohort dummies. We can then use the regression model that applies to the most recent available cohort for projecting forward (or even take account of trends in the cohort effects).[15]

The estimated human capital equation appears to perform well, with results that are broadly consistent with the existing literature. Focusing on Regression (1), we find a small negative effect of foreign experience. The coefficient on the years-since-migration variable shows that immigrant earnings grow at a real rate of roughly 2 percent per-year post-landing (after controlling for average economy-wide earnings). On language, we find that English is substantially more highly rewarded than French (approximately a 46 percent premium versus a 6 percent premium), no doubt reflecting the fact that a substantial majority of admitted immigrants move to English-speaking Canada.[16] The educational attainment variables broadly show the expected pattern, with the anomaly that Secondary shows slightly lower returns than Primary.[17] Interestingly, immigrants with a trade certificate have roughly equal earnings to those with some post-secondary attainment. The results show substantial earnings premiums are associated with higher

[15] Using just cohort dummies assumes that the coefficients on the human capital characteristics are stable over time. This assumption can be tested and, if necessary, relaxed using cohort-human capital variable interactions.

[16] An interesting extension would be to include interactions between the language variables and intended province of destination. This would, for example, allow us to explore the value of French for immigrants intending to reside in Quebec.

[17] The difference is not statistically significant in our base regression. The coefficient on Secondary becomes positive and statistically significant when we include all streams in the sample.

educational attainment. Compared to the base category, the premium for a bachelor's degree is approximately 19 percent higher than that for a diploma; a master's degree has a premium that is approximately 11 percent higher than that for bachelors; and a PhD has a premium that is approximately 29 percent higher than that for a masters. The macro variables have the expected signs. Most notably, the coefficient on the economy-wide average earnings variable is quantitatively large, with a 1 percent increase in economy-wide earnings associated with a 2.2 percent increase in immigrant earnings.[18]

Figure 1 displays the cohort effects (with the excluded cohort for 1980 equal to zero). The pattern of deteriorating cohort earnings holding human capital constant is consistent with previous findings—but the extent of the deterioration is dramatic.

Overall the regression explains just over 14 percent of the variation in log earnings. While this is broadly in line with the vast literature on human capital-based earnings regressions, it is an undeniably low number, suggesting that immigrant earnings performance is dominated by idiosyncratic factors. This is turn suggests fundamental limits to the points-based selection approach. In Section 5.3 we explore how the predictive power of the regression might be improved by adding additional observables. First, however, we show how a simple quasi-linear points system can be developed based on an illustrative regression.

5.2 An Illustrative Points System

The points system that is implied by Regression (1, OLS) is shown in Table 3. For this illustration, we assume that the discounted lifetime earnings threshold is set at $1,500,000 in constant 2004 dollars and the discount rate is set at 0.02. As described in Section 3, any applicant with a combination of characteristics that yields 100 points or more will be accepted (since 100 points or more means that lifetime predicted earnings that are at least $1,500,000). The underlying regression is somewhat more complicated than the simple example in Section 3, as it includes both cohort and macro effects in addition to measures of human capital at landing. Making the assumption that the most recent estimated cohort effect that is available (2003 in our sample) provides the best

[18] It is worth noting that this was a period of relatively low growth in economy-wide average earnings (just 0.5 percent over 1980 to 2004). Thus even with the high sensitivity to economy-wide earnings, there was still relatively low macro-related trend growth in immigrant earnings (approximately 1.1 percent).

indicator of future cohort effects, we add this effect to the regression constant to determine the constant for prediction equation for log earnings that underlies the points system. For the macro effects, we set the log of average earnings at its 2005 level ($47,800) and assume trend growth in earnings equal to the Bank of Canada's estimate for the underlying productivity growth rate (1.5 percent, Bank of Canada, 2006). We also assume that the unemployment rate is constant at its 2005 level (6.8 percent).

There are a number of notable features of the resulting point allocations. First, rather than being source of points, experience at landing actually attracts a small points penalty. Second, points for English exceed points for French by a factor of more than six. Third, while the highest point allocations are granted for to those with higher educational attainment, the holders of trade certificates also receive substantial points (comparable to someone with some post-secondary education). And fourth, and perhaps most surprisingly, age-at-landing has a dramatic impact on points. In our illustrative system, it is practically impossible for someone older than their mid-forties to meet the points threshold; on the other hand, it is hard for someone younger than their mid-twenties not to meet the threshold. The strong influence of age follows from the aggregation of lifetime earnings over potential years in the Canadian labour market (64 – Age at Landing). However, the relative impact of the age-at-landing variable can be attenuated by using a higher discount rate (matched by an appropriately lowered threshold), which effectively reduces the weight given to later years worked in Canada.

It is useful to examine a couple of examples to get a better feel for when someone succeeds or fails to make the threshold in this illustrative system. For our first example, we take a 37-year-old native English speaker with a master's degree and 12 years of experience.[19] The projected lifetime earnings of this individual is $1,700,082. Using an appropriately extended version of Equation (4), this projection can usefully be decomposed into the product of initial earnings ($38,529) and a factor that depends on potential years in Canada, the coefficient on the years since migration variable, the discount rate, and the product of the coefficient on the log of average annual earnings variable and the assumed growth rate for these earnings (44.124) – check this number. We call the latter number "adjusted potential years." The point allocations for this

[19] We assume they apply in 2005.

applicant are -1.1 for experience, 32.7 for language, 47.8 for educational attainment and 30.0 for age, for a total of 109.4 points. This applicant is thus accepted. For our second example, we take a 25-year-old native French speaker with a trade certificate and 5 years of experience. Initial projected earnings are $17,985 and adjusted potential years are 81.02, for total projected earnings of $1,487,065. This individual falls just below the projected earnings cutoff. Consistent with the earnings shortfall there is also a points shortfall: -0.5 for experience, 4.7 for language, 18.1 for educational attainment and 75.5 for age, for a total of 97.8. This applicant is thus rejected. However, if this individual was just one year younger with correspondingly one year less experience they would score 101.4 points (with projected lifetime earnings of $1,527,402) and would be accepted.

Even though this points system is just meant as an illustration, it is interesting to compare it the Canada's existing points grid. What are comparable are the relative points given for various pairs of characteristics rather than the absolute allocations of points. Most strikingly, under the present grid, 25 out of a maximum of 100 points are available for experience, while our findings suggest that no—or even slightly negative points—should be allocated for experience. For education, the same points (25) are allocated for a PhD and a Masters, which is not far above the points given for a two-year university degree (20). In contrast, our findings identify a much steeper educational attainment-points gradient. On age, the current grid calls for the maximum of age-related points to be given for applicants between 21 and 49 (10), with two-points per year penalties for each year above or below this range. Our findings identify both a larger relative weighting on age in general, and monotonically falling points with the actual age at landing. Because it is not limited to variables that are available in the IMDB, the current points system has the advantage of access to certain information not available to us. This information includes the presence of arranged employment, spouse's educational attainment, years of post-secondary study in Canada, and family relationships in Canada. However, these data could be collected in an expanded IMDB. We return to the benefits of expanding the range of individual data that is collected in the concluding comments below.

5.3 Extended Regressions

Taking the IMDB as given, we next explore how enriching the informational base for which points are given can lead to a better performing immigrant pool. We use the R^2 from the earnings regression as our measure of the predictive success of the selection system. In Appendix 1, we demonstrate how an increase in the R^2 leads to an outward shift in the selection frontier, so that the expected earnings of the admitted pool increases for any given number of admitted immigrants. We explore the addition of three types of information: non-linear terms for the foreign experience and years-since-migration variables; country-of-origin information; and intended-occupation dummies. To ensure valid comparisons, we limit our sample to observations for which all three forms of additional information are available. This causes the number of observations to drop from 313,631 to 258,175, and the number immigrants to drop from 50,160 to 43,218.

The results are recorded in Table 4. The first regression is our base regression (without the age-at-landing variable) estimated on the restricted sample. The results are very similar to those for the unrestricted sample. The next three regressions separately add the non-linear terms, the country-of-origin variables and the intended-occupation dummies to the base regression. The final regression adds all three forms of additional information simultaneously.

Regression (2) shows the effects of adding squared terms for Experience and Years Since Migration. The negative coefficient on Experience Squared shows that the proportionate earnings penalty on foreign experience rises, *cet. par.*, at an increasing rate with the extent of foreign experience. The coefficient on Years Since Migration Squared is also negative, indicating the growth rate of earnings tends to decline, *cet. par.*, with years in Canada. Indeed, on average earnings peak after just over 19 years. Overall, however, adding these squared terms adds minimally to the explanatory power of the regression, with the R^2 rising slightly from 0.1501 to 0.1541. Of course, adding two quadratic terms to the regression does not exhaust the potential for non-linearities. In particular, it is would be worthwhile to explore the explanatory power that comes from adding additional polynomial terms and interaction effects.

Regression (3) shows the effects of adding three country-of-origin variables: the log of real GDP per capita (adjusted for purchasing power parity); the log of distance from Canada, and the Gini coefficient (as a measure of source-country inequality). We hypothesized that the coefficient on the GDP per capita variable would be positive, reflecting such factors as better matched human capital when the immigrant is coming from another developed country with a more similar economy and also better source-country educational institutions. Our results support this hypothesis, with a 100 percent increase in source country GDP per capita resulting in a roughly 10 percent increase in earnings. We next hypothesized that an increase in the distance of the source-country from Canada would be positively associated with immigrant earnings. The reason is that the cost of immigrating will tend to rise with distance, and higher costs will increase the selection based on unobserved human capital characteristics. Put more simply, a higher cost of emigrating will tip the balance in favour of more skilled immigrants in any for any given levels of observed human capital. This hypothesis also receives support, with a 100 percent increase in distance leading to approximately an 11 percent increase in earnings. Finally, following Borjas (1987), we hypothesized that an increase in source-country inequality will tend to reduce immigrant earnings. The reason is that, where source-country inequality is high, more skilled individuals have a *relatively* stronger reason to stay at home, which should apply both between and within observed skill categories. Once again the hypothesis receives support, with a 10 point increase in the Gini coefficient being associated with an approximately 0.8 percent decrease in immigrant earnings. All told, the addition of the three country-of-origin variables only marginally increases the explanatory power of the regression, with the R^2 rising from 0.1501 to 0.1631.

Regression (4) shows the effects of introducing dummies for intended occupation. A dummy variable is introduced for each 2 digit National Occupation Classification (NOC) code with a number of additional classifications introduced by CIC. The introduction of the occupational dummies does lead to a more substantial improvement in the fit of the regression, with the R^2 rising from 0.1501 to 0.1936. It is worth emphasizing that this way of introducing occupation-specific information is quite different from the occupation-shortage approach that is, for example, an important aspect of the Australian points system. Under the latter approach, extra points are granted if there is deemed to be a shortage in the particular occupation. In contrast, our approach looks backwards to the earnings success of past immigrants destined for particular occupations. This has the disadvantage that it looks at past rather than present conditions in given occupations, but it has the advantage of focusing on how immigrants have actually done when destined for those occupations rather than economically dubious measures of shortage.[20] For example, there may be real shortages in certain health-related professions, but immigrants may face challenges in utilizing their human capital in those occupations because of difficulties getting their credentials recognized. Our approach has the merit of recognizing the de facto challenges in given occupations; although one could reasonably question the fairness of punishing future applicants because of inefficient credential recognition in the past.

Table 5 records the occupation effects (measured in log points), which are ordered by size of effect. The omitted category is the CIC category of "new worker" (NOCD9914). A small number of the CIC categories did not contain any observations in our sample and are not listed in the table. The first column of the table records the share of our sample destined for the occupation. Clearly, some of the shares are quite small, so that the estimated effects should be treated with some caution. However, the pattern of effects looks largely plausible, with immigrants intending to enter senior management earning the largest premium over new workers, while those intending to be homemakers earning the lowest. By far the largest category is NOCD21 "professional occupations in natural and applied sciences," with almost a quarter of the sample. Workers intending to enter this occupational category earn a premium over new workers of approximately 50 percent.

Regression (5) finally adds all three forms of information in a single regression. Overall, the fit of the regression increases by more than one-third. But the percentage of variation in log earnings explained is still low at just over 20 percent**.** While we obviously have not exhausted the types of information that could be used in the underlying prediction regression, the evident difficulty of predicting who will succeed economically based on observed human capital characteristics at landing leads one to ask if there is a better way to select economic immigrants. We take up this question in the concluding section.

---

[20] Often it seems that the term "shortage" is used where there is upward pressure on wages.

6. Concluding Comments

Is there a better way to select economic immigrants? The limited predictive power of the models we have explored certainly motivates a search for alternatives. A leading contender is U.S.-style employer-driven selection. Employers are obviously motivated to work hard to identify talented individuals. They can also utilize a richer informational base: Where did they get their education? How well do they speak the language? How likely is it that their recommenders value their reputations for honest evaluations? Put simply, employers are well-placed to be "experts" when it comes to predicting who will be successful on the job. While granting that employers often have information that cannot easily be integrated into a points system, it is important to recognize that there is a vast literature on the superiority of actuarial/statistical-based over clinical/expert-based judgment across a range of settings (see Grove et al., 2000, for a meta-analysis).[21] This edge is often present even when the clinician has information that is not available for the statistical analysis (Grove and Meehl, 1996). The typical, and for many surprising, superiority of the actuarial approach stems from a combination of its edge in solving the complex problem of appropriately weighting disparate pieces of information and its avoidance of biases that afflict subjective judgment.[22]

Our prediction, for what its worth, is that employer evaluations will have a critical role to play. But the most effective feasible selection system is likely to be one that integrates the informational value of employer assessments into actuarial-based predictions. The evidence from other areas suggests that this should be done, not by allowing employers to selectively override the points system, but by turning the employer information sources of points. This could be done, for example, by giving points based on the existence of job offers, salary offers, past home-country salaries, and so on.

---

[21] Paul Meehl, a pioneer in the making of these comparisons, sums up the literature as follows:

> There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing over 100 investigations, predicting everything from the outcome of football games to the diagnosis of liver disease, and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion (Quoted in, Ayres, 2007, p. 127).

[22] See Ayers (2007, Chapter 5) for an accessible recent discussion.

The Australian experience also suggests that better selections can be made by improving the informational quality of the type of variables that are currently used: better language ability testing through formal language tests, say; or better measures of educational attainment by making use of objective rankings of educational institutions. Since the actuarial method looks backwards to determine the optimal weights to place the various pieces of predictive information, it is important to begin collecting the more fine-grained information as soon as possible. For the IMBD, this means adding new variables to the "tombstone data" part of the database. This additional data could be collected on a random sample of admitted immigrants until it has proven its value for prediction. Notwithstanding the challenges of predicting immigrant success, if the policy objective is to bring in individuals who will succeed economically, an actuarially-based selection system operating on an appropriately rich informational base will be hard to trump.

Appendix 1. The Selection Frontier

The selection frontier shows the menu of choices available to the policy maker given the human-capital based earnings regression used to predict immigrant success. More precisely, the frontier shows the tradeoff between the expected earnings of the admitted pool and the number of immigrants admitted. Choosing a point on the frontier is equivalent to choosing a predicted-earnings threshold. The combination of the threshold and equation (6) then completely defines the points system.

To identify the frontier, we assume that the actual host-country earnings of the applicant pool (i.e. the pool of individuals who desire to emigrate to the potential host-country) are log-normally distributed. (For simplicity, we restrict attention to the case of a single-period immigrant horizon in the host-country.) In McHale and Rogers (2007), we show that this implies that actual earnings and predicted earning of potential immigrants are joint-normally distributed with a correlation coefficient equal to the coefficient of determination (R^2) in the prediction regression. The admitted pool is then an incidental truncation of the applicant pool, with selection into the applicant pool based on a comparison on predicted earnings and the predicted-earnings threshold. Letting $\hat{y}^*$ be the single-year predicted earnings threshold (and dropping i subscripts for notational convenience), the expected earnings of the admitted immigrant pool is given by,[23]

$$\text{(A.1)} \quad E[Y \mid \hat{Y} \geq \hat{Y}^*] = E[Y \mid \hat{y} \geq \hat{y}^*] = \overline{Y}\left[\frac{1-\Phi(z_{\hat{y}}(\hat{y}^*)-\sigma_{\hat{y}}))}{1-\Phi(z_{\hat{y}}(\hat{y}^*))}\right],$$

where $z_{\hat{y}}(k) = (\dfrac{k-\mu_{\hat{y}}}{\sigma_{\hat{y}}})$ and $\overline{Y}$ is the (arithmetic) mean earnings of the applicant pool.

The share of the application pool that is admitted is also a function of the predicted-earnings threshold,

$$\text{(A.2)} \quad p = 1-\Phi(z_{\hat{y}}(\hat{y}^*)).$$

[23] See McHale and Rogers (2007) for details.

Combining equations (7) and (8) we can eliminate $\hat{y}^*$ and obtain an equation for the selection frontier,

$$\text{(A.3)} \quad E[Y \mid p] = \frac{\overline{Y}}{p}(1 - \Phi(\Phi^{-1}(1-p) - (\sigma_y{}^2 - \sigma_u{}^2)^{\frac{1}{2}}))\,.$$

The selection frontier is shifted upwards—and thus the tradeoff facing the policy maker improved—by an increase in the mean earnings of the applicant pool ($\overline{Y}$), an increase in the variance of log earnings in the applicant pool (σ_y^2), and a reduction in the variance of the prediction error for log earnings (σ_u^2).

The selection frontier and the associated predicted-earnings thresholds are shown in Figure 1 for the case where $\overline{Y}$ = \$65,000, σ_y^2 = 0.4, and σ_u^2 = 0.32. Note that the latter two parameters imply that the R^2 from the earnings regression is equal to 0.2. Figure 2 shows how improvements in the predictive accuracy of the of human-capital earnings regression (here captured by increases in the R^2) leads to increases in expected earnings for any given share of the applicant pool that is admitted. The selection frontier thus allows us to explore how improvements in the ability to predict immigrant labour market leads to changes in the average "quality" of selected immigrants.

Figure A.1

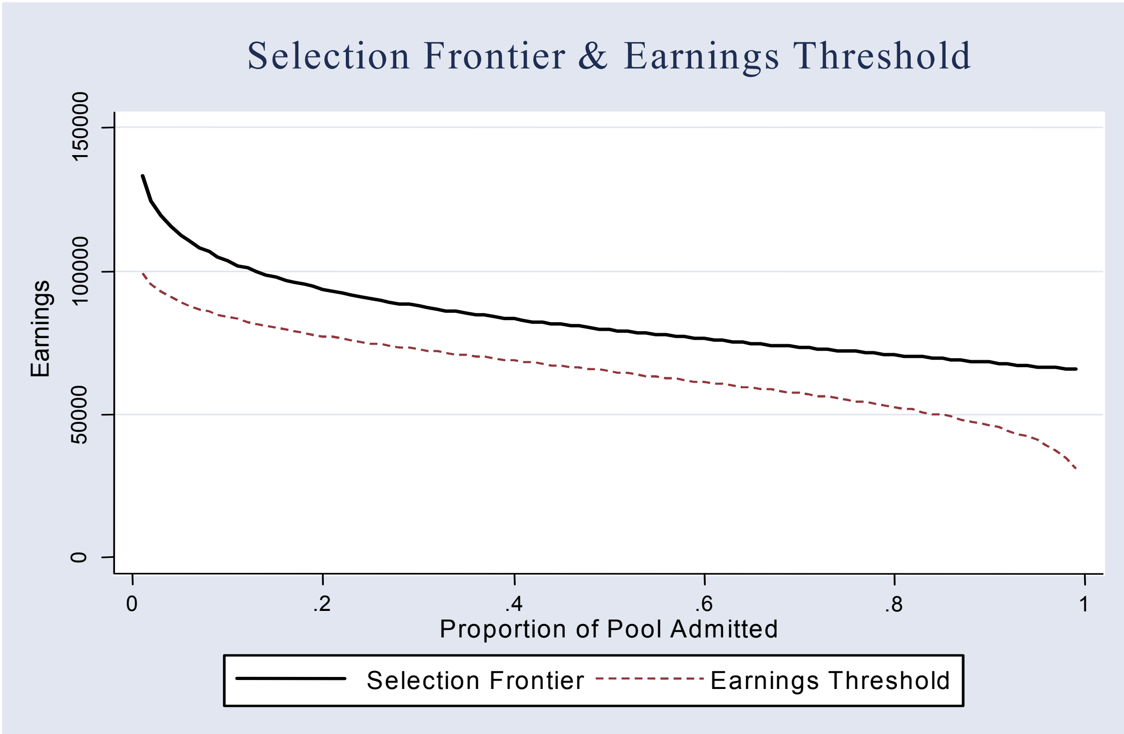
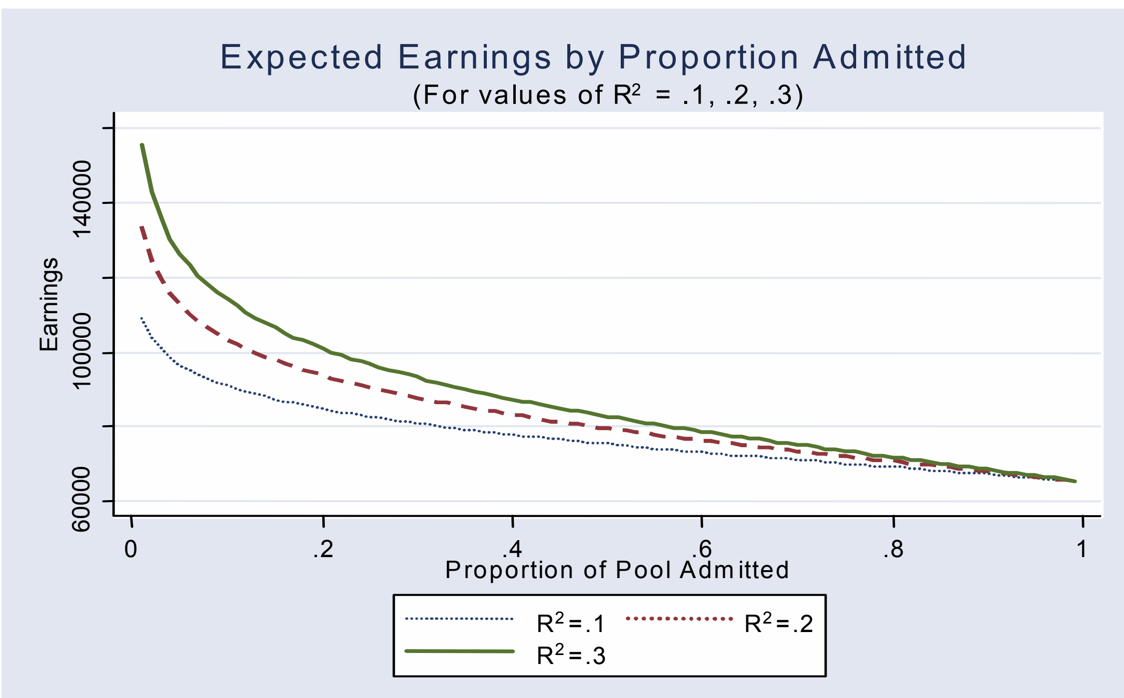


Figure A.2



References

Abbot, Michael (2003), "The IMDB: A User's Overview of the Immigration Database," in *Canadian Immigration Policy for the 21st Century*, Charles Beach, Alan Green, and Jeffrey Reitz, editors, Montreal and Kingston: McGill-Queen's University Press, pp. 322.

Abbot, Michael, and Charles Beach (1993), "Immigrant Earnings Differentials and Birth-Year Effects for Men in Canada: Post-War-1972," *Canadian Journal of Economics*, Vol 26, Issue 3, pp. 505-524.

Alborim, Naomi, Ross Finnie, and Ronald Meng (2005), "The Discounting of Immigrants' Skills in Canada: Evidence and Policy Recommendations," *IRPP Choices,* 11(2), Montreal: Institute for Research in Public Policy.

Antecol, Heather, Deborah Cobb-Clark, and Stephen Trejo (2001), "Immigration Policy and the Skills of Immigrants to Australia, Canada, and the United States," IZA Discussion Paper No. 363.

Aydemir, Abdurrahman (2002), "Effects of Selection Criteria and Economic Opportunities on the Characteristics of Immigrants," Statistics Canada, Research Paper Series.

Aydemir, Abdurrahman, and Mikal Skuterud (2005), "Explaining the Deteriorating Entry Earnings of Canada's Immigrant Cohorts, 1966-2000," *Canadian Journal of Economics*, Vol 38, No. 2, pp 641-672.

Ayres, Ian (2007), *Super Crunchers: Why Thinking-by-Numbers is the New Way to be Smart*, New York: Bantham Books.

Baker, Michael, and Dwayne Benjamin (1994), "The Performance of Immigrants in the Canadian Labor Market," *Journal of Labor Economics,* Vol. 12, No. 3, pp. 369-405.

Bank of Canada (2006), *Monetary Policy Report*, October, Ottawa: Bank of Canada.

Beach, Charles (2006), "Employment-Based Permanent Immigration: Examining the Value of a Skills-Based Point System," Statement before the U.S. Senate Committee on Health, Education, Labor, and Pensions, September 14.

Beach, Charles, Alan Green, and Christopher Worsick (2006), "Impacts of the Point System and Immigration Policy Levers on Skill Characteristics of Canadian Immigrants," Working paper, Queen's University and Carleton University.

Bloom, David, Gilles Grenier, and Morley Gunderson (1995), "The Changing Labour Market Position of Canadian Immigrants," *Canadian Journal of Economics*, Vol 28, Issue 4b, pp. 987-1005.

Borjas, George (1985), “Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants,” *Journal of Labor Economics*, Vol. 3, No. 4, pp. 463-489.

Borjas, George (1987), “Self Selection and the Earnings of Immigrants,” *American Economic Review*, 77(4), pp. 531-553.

Borjas, George (1995), “The Economic Benefits of Immigration,” *Journal of Economic Perspectives*, 19(2), pp. 2-22.

Chiswick, Barry (1978), “The Effect of Americanization on the Earnings of Foreign-Born Men,” *Journal of Political Economy*, 86, pp. 897-921.

Dawes, Robyn, David Faust, and Paul Meehl (1989), “Clinical Versus Actuarial Judgment,” *Science*, Vol 243, No. 4899, pp. 1668-1674.

Duleep, Harriet Orcutt, and Mark Regrets (2002), “The Elusive Concept of Immigrant Quality: Evidence form 1970-1990,” IZA Working Paper No. 631, November.

Ferrer, Ana, and W. Craig Riddell (2004), “Education, Credentials and Immigrant Earnings,” working paper, University of British Columbia, August.

Ferrer, Ana, David Green, and W. Craig Riddell (2005), “The Effect of Literacy on Immigrant Earnings,” working paper, University of British Columbia, August.

Frenette, Marc, and Rene Morissette (2003), “Will They Ever Converge? Earnings of Immigrant and Canadian-Born Workers Over the Last Two Decades,” Statistics Canada, Analytical Studies Branch Research Paper Series.

Friedberg, Rachel (2002), “You Can’t Take It with You? Immigrant Assimilation and the Portability of Human Capital,” *Journal of Labor Economics*, 18(2), pp 221-251.

Glenn, Norval (2005), *Cohort Analysis,* Second Edition, Series on Quantitative Applications in the Social Sciences, London: Sage Publications

Grant, Mary (1999), “Evidence of New Immigrant Assimilation in Canada,” *Canadian Journal of Economics*, Vol. 32, No. 4, pp. 930-955.

Green, David, and Christopher Worswick (2004), “Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Marco Effects,” Working Paper University of British Columbia and Carleton University.

Grove, Willian, and Paul Meehl (1996), “Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy,” *Psychology, Public Policy, and Law*, 2(2), pp. 293-323.

Grove, William, David Zald, Boyd Lebow, Beth Snitz, and Chad Nelson (2000), "Clinical Versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 12(1), pp. 19-30.

Hall, Robert (1971), "The Measurement of Quality Change from Vintage Price Data," in Zvi Grilliches (ed.), *Price Indexes and Quality Change*, Cambridge: Harvard University Press, pp. 240-271.

Hawthorne, Lesleyanne (2005), "Picking Winners": The Recent Transformation of Austrialia's Skilled Migration Policy," *International Migration Review*," Vol. 39, No. 3, pp. 663-696.

Jasso, Guillermina, and Mark Rosenzweig (1995), "Do Immigrants Screened for Skills do Better than Family Reunification Immigrants?" *International Migration Review*," Vol. 29, No. 1, pp. 85-111.

Jasso, Guillermina, and Mark Rosenzweig (2005), "Selection Criteria and the Skill Composition of Immigrants: A Comparative Analysis of Australian and U.S. Employment Immigration," forthcoming in Jagdish Bhagwati and Gordon Hanson, editors, *Skilled Migration Today: Prospects, Problems, and Policies*.

Mason, Oppenheim Karen, William M. Mason, H.H. Winsborough, and W. Kenneth Poole (1973), "Some Methodological Issues in Cohort Analysis of Archival Data," *American Sociological Review*, 38(2), pp. 242-258.

McHale, John, and Keith Rogers (2007), "Optimal Design of an Immigration Points System," Working paper, Queen's University.

Meehl, Paul (1954), *Clinical Versus Statistical Prediction*, Minneapolis: University of Minnesota Press.

Picot, Garnett, and Arthur Sweetman (2005), "The Deteriorating Economic Welfare of Immigrants and Possible Causes: Update 2005," Statistics Canada, Analytical Studies Branch Research Paper Series.

Picot, Garnett, Feng Hou, and Simon Coulombe (2007), "Chronic Low Income and Low-Income Dynamics Among Recent Immigrants," Statistics Canada, Analytical Studies Branch Research Paper Series.

Schaafsma, Joseph, and Arthur Sweetman (2001), "Immigrant Earnings: Age at Immigration Matters," *Canadian Journal of Economics*, 34(4), pp. 1066-1099.

Sweetman, Arthur (2004), "Immigrant Source Country Educational Quality and Canadian Labour Market Outcomes," Statistics Canada, Analytical Studies Branch Research Paper Series, No. 234, December.

United Kingdom Home Office (2006), "*A Points-Based System: Making Migration Work for Britain*," report present to Parliament by the Secretary of State for the Home Department.

Figure 1. Estimated Cohort Year Effects (1980 = 0), 1980 - 2003

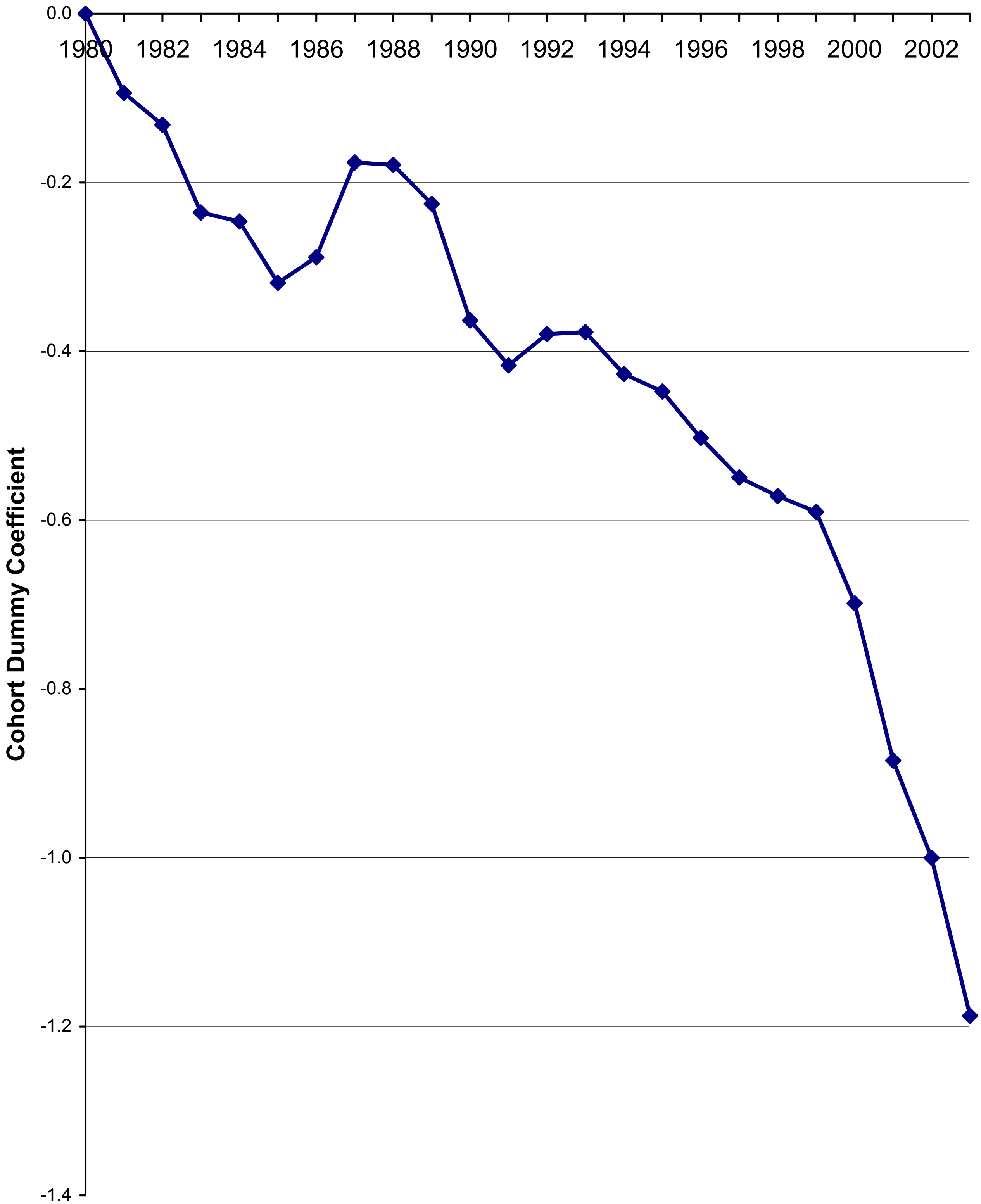


Table 1. Summary Statistics

	Obs.	Median	Mean	S.D
Individual Variables				
Log Annual Earnings (constant 2005 Dollars)	314,892	10.44	10.29	0.95
Landing Year	314,892	1991	1991	6.49
Tax Year	314,892	1998	1996	5.97
Age at Landing	314,892	32	33	7.14
Years Since Migration	314,892	5	6	3.80
Years of Schooling	313,631	15	14	3.83
Experience	313,631	12	14	7.70
English Language (mother tounge)	314,892	0	0.21	0.41
French Language (mother tounge)	314,892	0	0.05	0.21
Primary	314,892	0	0.09	0.28
Secondary	314,892	0	0.14	0.34
Some Post-Secondary	314,892	0	0.07	0.26
Trade Certificate	314,892	0	0.14	0.34
Diploma	314,892	0	0.11	0.31
Bachelors	314,892	0	0.32	0.47
Masters	314,892	0	0.10	0.29
PhD	314,892	0	0.04	0.19
Macro Variables				
Unemployment Rate	314,892	0.078	0.086	0.01
Log Average Annual Earnings (constant dollars)	314,892	10.75	10.73	0.04
Country-of-Origin Variables				
Log GDP-Per-Capita	284,866	8.36	8.62	1.01
Log Distance (from Canada)	299,074	8.66	8.52	0.36
Gini	279,736	39.33	39.77	7.30

Sample includes Skilled Workers, Principal Applicants (IMCAT categories 7 & 8) with earnings > $1,000

Table 2. Basic Log Earnings Regressions
Skilled Workers, Principal Applicants

Dependent Variable = Log Annual Earnings

	(1)		(2)	
	OLS	Random Effects	OLS	Random Effects
Age at Landing	. . .	. . .	0.0201 *	0.0170 *
			(0.0020)	(0.0021)
Years Since Migration	0.0202 *	0.0142 *	0.0202 *	0.0142 *
	(0.0009)	(0.0006)	(0.009)	(0.0006)
Experience	-0.0013 **	-0.0023 *	-0.0203 *	-0.0184 *
	(0.0005)	(0.0005)	(0.0019)	(0.0021)
English Language	0.4366 *	0.5268 *	0.4292 *	0.5197 *
	(0.0094)	(0.0105)	(0.0094)	(0.0105)
French Language	0.0621 *	0.1403 *	0.0655 *	0.1422 *
	(0.0165)	(0.0174)	(0.0164)	(0.0174)
Secondary	-0.0026	0.0009	-0.0837	-0.0666 *
	(0.0154)	(0.0120)	(0.0172)	(0.0216)
Some Post-Secondary	0.2331 *	0.2150 *	0.0858 *	0.0910 *
	(0.0189)	(0.0223)	(0.0236)	(0.0270)
Trade Certificate	0.2421 *	0.2212 *	0.1319 *	0.1276 *
	(0.0161)	(0.0202)	(0.0192)	(0.0232)
Diploma	0.3315 *	0.3161 *	0.1862 *	0.1940 *
	(0.0170)	(0.0208)	(0.0219)	(0.0257)
Bachelors	0.5291 *	0.5037 *	0.3463 *	0.3515 *
	(0.0154)	(0.0188)	(0.0232)	(0.0266)
Masters	0.6383 *	0.5846 *	0.4188 *	0.4020 *
	(0.0182)	(0.0209)	(0.0279)	(0.0380)
PhD	0.9294 *	0.8862 *	0.6602 *	0.6606 *
	(0.0218)	(0.0259)	(0.0340)	(0.0380)
Unemployment Rate	-3.3675 *	-3.8351 *	-3.3743 *	-3.8357 *
	(0.1500)	(0.1067)	(0.1500)	(0.1067)
Log Average Annual Earnings	2.2447 *	2.6093 *	2.2454 *	2.6095 *
	(0.1129)	(0.0832)	(0.1129)	(0.0832)
Cohort Year Dummies	Yes	Yes	Yes	Yes
R Squared	0.1410	0.1386	0.1423	0.1400
Within	. . .	0.0607	. . .	0.0607
Between	. . .	0.1352	. . .	0.1362
Root Mean Square Error	0.8770	. . .	0.0876	. . .
Observations	313,631	313,631	313,631	313,631
Individuals	50,610	50,610	50,610	50,610

Standard errors are in parentheses; OLS standard errors are robust to individual-level clustering.
* = signficance at 1% level; ** = significance at 5% level.
Sample includes Skilled Workers, Principal Applicants (IMCAT categories 7 & 8) with annual earnings > $1,000

Table 3. Ilustrative Points System

Points threshold = 100; Points allocations based on Regression 1 (OLS), Table 2

Experience (per year)	-0.1
Language	
English	32.7
French	4.7
Eduational Attainment	
Secondary	-0.2
Some Post-Secondary	17.5
Trade Certificate	18.1
Diploma	24.8
Bachelors	39.6
Masters	47.8
PhD	69.6

Age at Landing		*Age at Landing (continued)*	
63	-252.4	40	16.8
62	-199.2	39	21.3
61	-167.5	38	25.7
60	-144.7	37	30.0
59	-126.7	36	34.2
58	-111.7	35	38.3
57	-98.9	34	42.3
56	-87.5	33	46.2
55	-77.4	32	50.1
54	-68.2	31	53.9
53	-59.7	30	57.6
52	-51.8	29	61.3
51	-44.5	28	64.9
50	-37.5	27	68.5
49	-31.0	26	72.0
48	-24.8	25	75.5
47	-18.9	24	78.9
46	-13.2	23	82.3
45	-7.8	22	85.7
44	-2.5	21	89.0
43	2.6	20	92.3
42	7.5	19	95.6
41	12.2	18	98.8

Assumptions:

1. Lifetime earnings threshold (constant 2004 dollars)	1,500,000
2. Assumed trend growth in average annual earnings	1.50%
3. Average Earnings in 2003 (full year, full time)	48,700
4. Unemployment rate in 2003	6.8%
5. Discount rate (δ)	0.02

Table 4. Extended OLS Regressions: Skilled Workers, Principal Applicants
Non-linearities, Country of Origin, Intended Occupations

Dependent Variable = Log Annual Earnings

	(1)	(2)	(3)	(4)	(5)
Years Since Migration	0.0183 *	0.0763 *	0.0190 *	0.0191 *	0.0784 *
	(0.0010)	(0.0020)	(0.0010)	(0.0010)	(0.0019)
Years Since Migration Squared	. . .	-0.0040 *	. . .	. . .	-0.0041 *
		(0.0001)			(0.0001)
Experience	-0.0018 *	0.0044 *	-0.0013 *	-0.0042 *	-0.0006 *
	(0.0006)	(0.0016)	(0.0010)	(0.0005)	(0.0016)
Experience Squared	. . .	-0.0002 *	. . .	. . .	-0.0001 *
		(0.0000)			(0.0000)
English Language	0.4774 *	0.4805 *	0.4242 *	0.4418 *	0.4166 *
	(0.0106)	(0.0106)	(0.0118)	(0.0101)	(0.0112)
French Language	0.0525 *	0.0587 *	-0.0654 *	0.0987 *	0.0222
	(0.0177)	(0.0178)	(0.0185)	(0.0173)	(0.0181)
Secondary	-0.0009	-0.0066	-0.0275	0.0212	-0.0067
	(0.0181)	(0.0181)	(0.0174)	(0.0171)	(0.0170)
Some Post-Secondary	0.2700 *	0.2637 *	0.2378 *	0.2081 *	0.1863 *
	(0.0218)	(0.0218)	(0.0213)	(0.0210)	(0.0209)
Trade Certificate	0.2523 *	0.2437 *	0.1855 *	0.1623 *	0.1184 *
	(0.0188)	(0.0187)	(0.0182)	(0.0178)	(0.0175)
Diploma	0.3610 *	0.3532 *	0.3051 *	0.2433 *	0.2101 *
	(0.0197)	(0.0198)	(0.0192)	(0.0195)	(0.0194)
Bachelors	0.5395 *	0.5326 *	0.5029 *	0.3526 *	0.3376 *
	(0.0179)	(0.0180)	(0..0174)	(0.0186)	(0.0185)
Masters	0.6469 *	0.6418 *	0.6298 *	0.4394 *	0.4412 *
	(0.0206)	(0..0206)	(0.0201)	(0.0126)	(0.0215)
PhD	0.9345 *	0.9295 *	0.9067 *	0.7185 *	0.7142 *
	(0.0239)	(0..0239)	(0.0233)	(0.0258)	(0.0255)
Unemployment Rate	-3.0114 *	-2.7417 *	-3.0245 *	-3.0247 *	-2.7636 *
	(0.1677)	(0.1673)	(0.1667)	(0.1648)	(0.1641)
Log Average Annual Earnings	2.4380 *	2.3702 *	2.3806 *	2.4184 *	2.3143 *
	(0.1238)	(0.1231)	(0..1236)	(0.1221)	(0.1213)
Log GDP Per Capita	. . . .	. . .	0.0971 *	. . .	0.0782 *
			(0.0048)		(0.0047)
Log Distance	. . . .	. . .	0.1103 *	. . .	0.1018 *
			(0.0146)		(0.0141)
Gini	. . . .	. . .	0.0083 *	. . .	-0.0062 *
			(0.0005)		(0.0005)
Cohort Dummies	Yes	Yes	Yes	Yes	Yes
Occupational Dummies	No	No	No	Yes	Yes
R Squared	0.1501	0.1541	0.1631	0.1936	0.2040
Root MSE	0.8768	0.8748	0.8708	0.8542	0.8487
Observations	258,175	258,175	258,175	258,175	258,175
Immigrants	43,218	43,218	43,218	43,218	43,218

Standard errors are in parentheses; OLS standard errors are robust to individual-level clustering.
* = signficance at 1% level; ** = significance at 5% level.
Sample includes Skilled Workers, Principal Applicants (IMCAT categories 7 & 8) with annual earnings > $1,000

Table 5. Estimates of Occupation Effects

Based on Regression (4), Table 4

		Share	Cofficient	Stand. Error
Omitted Occupational Category, NOCD9914, New worker (CIC)		6.036%	0.0000	. . . .
NOCD				
00	Senior management occupations	0.367%	1.2364	0.0814
02	Middle and other management occupations	0.262%	0.9628	0.0770
09	Middle and other management occupations	0.463%	0.6797	0.0712
08	Middle and other management occupations	0.076%	0.6221	0.1628
01	Middle and other management occupations	1.182%	0.5535	0.0405
92	Processing, manufacturing and utilities supervisors and skilled oper.	0.227%	0.5198	0.0719
21	Professional occupations in natural and applied sciences	23.524%	0.4911	0.0176
9990	Software pilot (CIC)	0.003%	0.4769	0.3114
72	Trades and skilled transport and equipment operators	5.876%	0.4525	0.0198
07	Middle and other management occupations	0.370%	0.4368	0.0575
11	Professional occupations in business and finance	4.258%	0.4165	0.0222
22	Technical occupations related to natural and applied sciences	7.024%	0.3833	0.0195
31	Professional occupations in health	3.169%	0.3356	0.0252
95	Processing and manufacturing machine operators and assemblers	0.863%	0.3339	0.0369
65		0.037%	0.3279	0.1421
73	Trades and skilled transport and equipment operators	6.046%	0.3125	0.0197
04	Middle and other management occupations	0.009%	0.3010	0.3023
03	Middle and other management occupations	0.100%	0.2935	0.1773
05	Middle and other management occupations	0.235%	0.2922	0.0912
06	Middle and other management occupations	2.163%	0.2388	0.0310
41	Prof. occs in social science, education, gov. services and religion	4.552%	0.2372	0.0228
96	Labourers in processing, manufacturing and utilities	0.626%	0.2006	0.0435
32	Techical and skilled occupations in health	1.359%	0.1786	0.0294
74	Intermediate occs in trans., equip. operation, install. and main.	0.684%	0.1782	0.0458
76	Trades helpers, construction labourers and related occupations	0.731%	0.1650	0.0375
94	Processing and manufacturing machine operators and assemblers	1.725%	0.1443	0.0277
9999	Open employment authorization (CIC)	0.015%	0.1302	0.1906
82	Skilled occupations in primary industry	0.335%	0.1290	0.0684
62	Skilled sales and service occupations	5.939%	0.0824	0.0194
12	Skilled administrative and business occupations	7.058%	0.0813	0.0188
84	Intermediate occupations in primary industry	0.820%	0.0721	0.0355
34	Assisting occupations in support of health services	0.283%	0.0480	0.0669
14	Clerical occupations	2.464%	0.0444	0.0244
52	Technical and skilled occupations in art, culture, and recreation	1.122%	0.0094	0.0342
64	Intermediate sales and services occupations	4.287%	-0.0068	0.2141
42	Paraprofessional occs in law, social services, education and religion	0.502%	-0.0219	0.0425
9911	Student (CIC)	1.309%	-0.0722	0.0335
51	Professional occupations in art and culture	1.473%	-0.1043	0.0326
66	Elemental sales and service occupations	2.017%	-0.1756	0.0249
9992	Retired (CIC)	0.039%	-0.3035	0.1829
9980	Other non-worker (CIC)	0.067%	-0.3679	0.1242
9970	Homemaker (CIC)	0.261%	-0.4477	0.0700

Occupations are ordered by size of occupation effect.
Occupational codes were supplied by CIC; they include aggregated NOC codes plus special CIC categories.
Occupational categories without observations in our sample are not listed.